

Through a Lens, Brightly:

How the worldview of incredible fiction describes the
spirit of the age

*Applying digital tools to analyse a
fictional genre*

Ben Menadue, James Cook University CASE
Supervisors: Richard Lansdown, Komla Tsey, Susan Jacups
Digital Humanities 3rd November 2016



2001

a space odyssey
a novel by
Arthur C. Clarke



based on the screenplay by
Stanley Kubrick & Arthur C. Clarke

Science Fiction

🎬 unbounded

🎬 inclusive

🎬 immediate

“the future now”

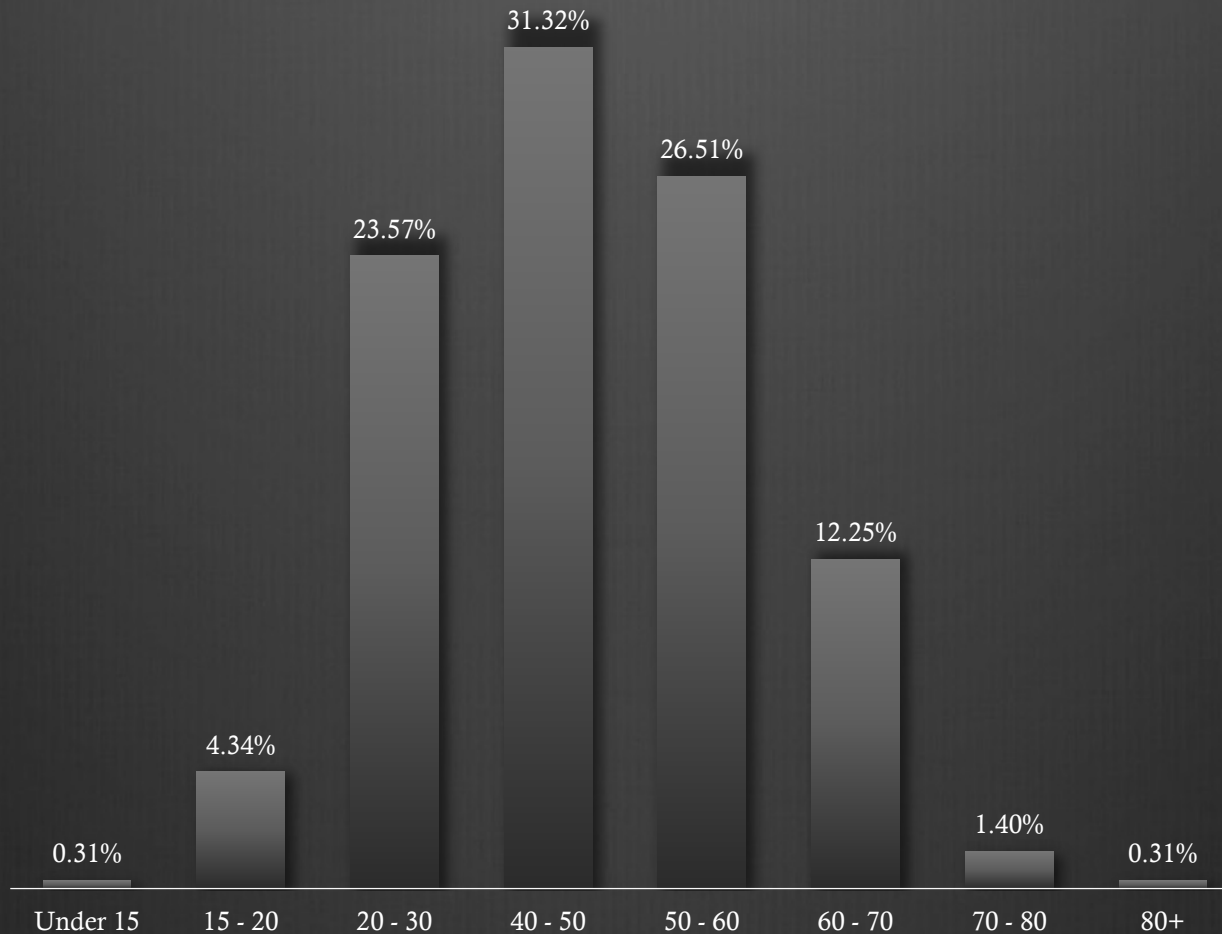
(Ballard, 1986)

Why science fiction?

- ❉ Popular fiction reflects human interests – science fiction has become ‘normal’ fiction in demographic appeal
- ❉ A lack of content limitations in science fiction encourages exploratory thematic and mythological content as well as encouraging writing that mirrors “the future now”
- ❉ Other forms of fiction are more constrained by genre characteristics
- ❉ Magazine science fiction is rapidly produced, and ephemeral – a form of literature closely linked to contemporary society
- ❉ It includes editorial, reader feedback and advertising content that provides additional contextual information

A standard demographic (compared to US census data)

science fiction consumer age profiles
current online survey built with Google Forms (n=865)





There's a lot of it to work with...

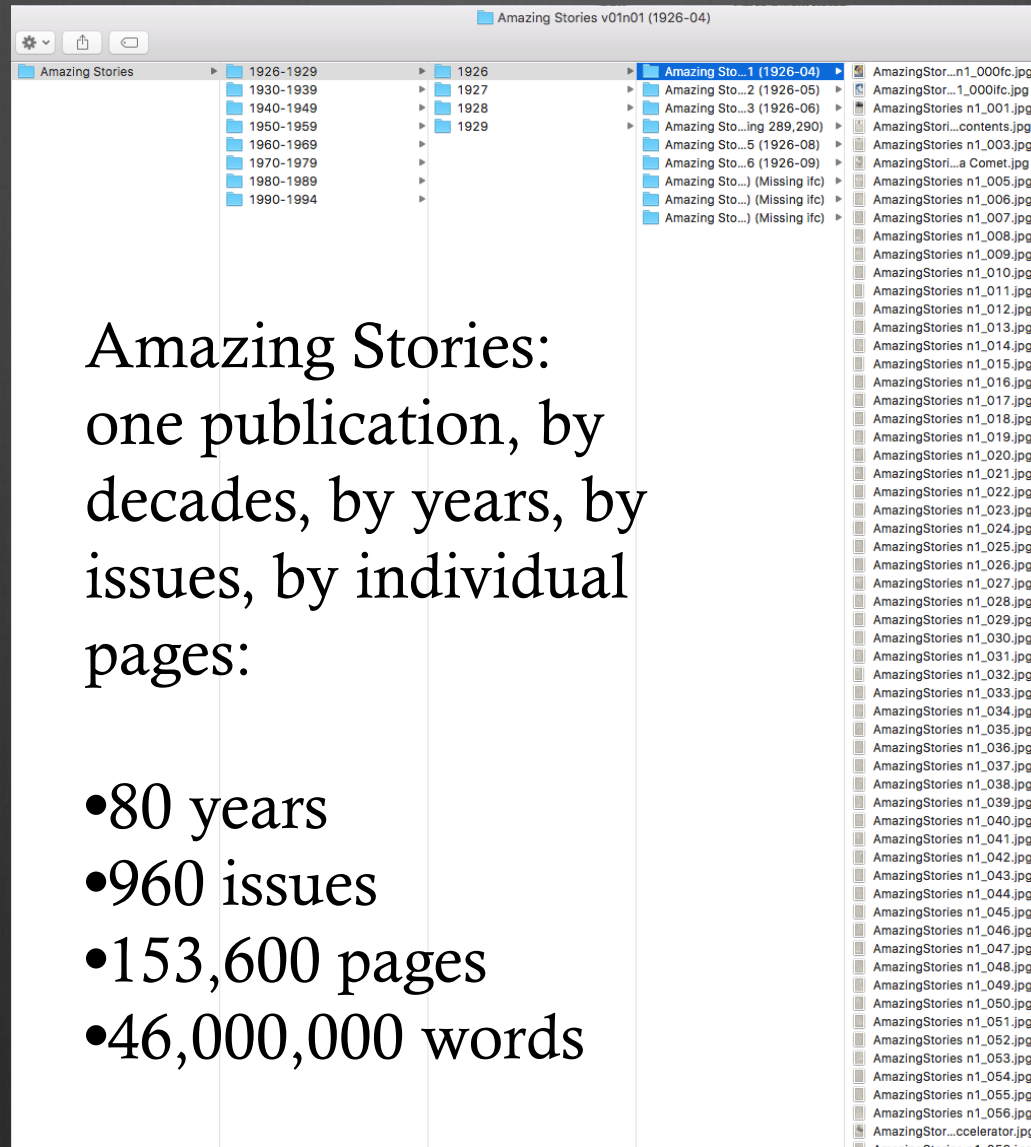
Theoretical approach

- ⊗ Synthesising quantitative and qualitative methods offers the potential to find results that demonstrate improved statistical robustness over purely qualitative approaches
- ⊗ Computer technology now exists both for quantitative analysis and the coding of qualitative data, which makes a large-scale approach feasible (and Franco Moretti goes even bigger, and broader)
- ⊗ It was impossible until the latter part of the 1980s to carry out meaningful computational text analysis at all
- ⊗ Today manually attempting the scale of my work is simply unfeasible (reading one text a day would take 12 years)

Digital data collection

- ⦿ Primary literature analysis
 - ⦿ Quick
 - ⦿ Methodical
 - ⦿ Quantitative
 - ⦿ Inexpensive (if you already have the material...)
- ⦿ Online surveys
 - ⦿ Cheap and surprisingly easy
 - ⦿ Qualitative and quantitative (Likert scale and free text answers)

Methodology “zoomed out”: the scale of the digital approach



Amazing Stories:
one publication, by
decades, by years, by
issues, by individual
pages:

- 80 years
- 960 issues
- 153,600 pages
- 46,000,000 words

Automation (Mac)

Adobe pdfs created from folders of page images

Library

- Calendar
- Contacts
- Developer
- Documents
- Files & Folders
- Fonts
- Internet
- Mail
- Movies
- Music
- PDFs
- Photos
- Presentations
- System
- Text
- Utilities
- Other
- Most Used
- Recently Added

Activate Fonts

This action activates the fonts passed from the previous action.

Input: Font Book typeface
Result: Font Book typeface

Repeat for each subfolder found

Results **Options**

Dispense Items Incrementally

Results **Options**

Set Value of Variable

Variable: **Filename**

Results **Options**

New PDF from Images

Save Output to: **Desktop**

Output File Name: **New PDF from Images Output** ☐ Replace Existing Files

☒ Size Each Page to Fit
☐ Make All Pages the Same Size
☐ Make All Pages this Size: width: height: points

Results **Options**

Rename Finder Items: Name Single Item

Name Single Item ☒

Name: **Basename only** to **FolderName**

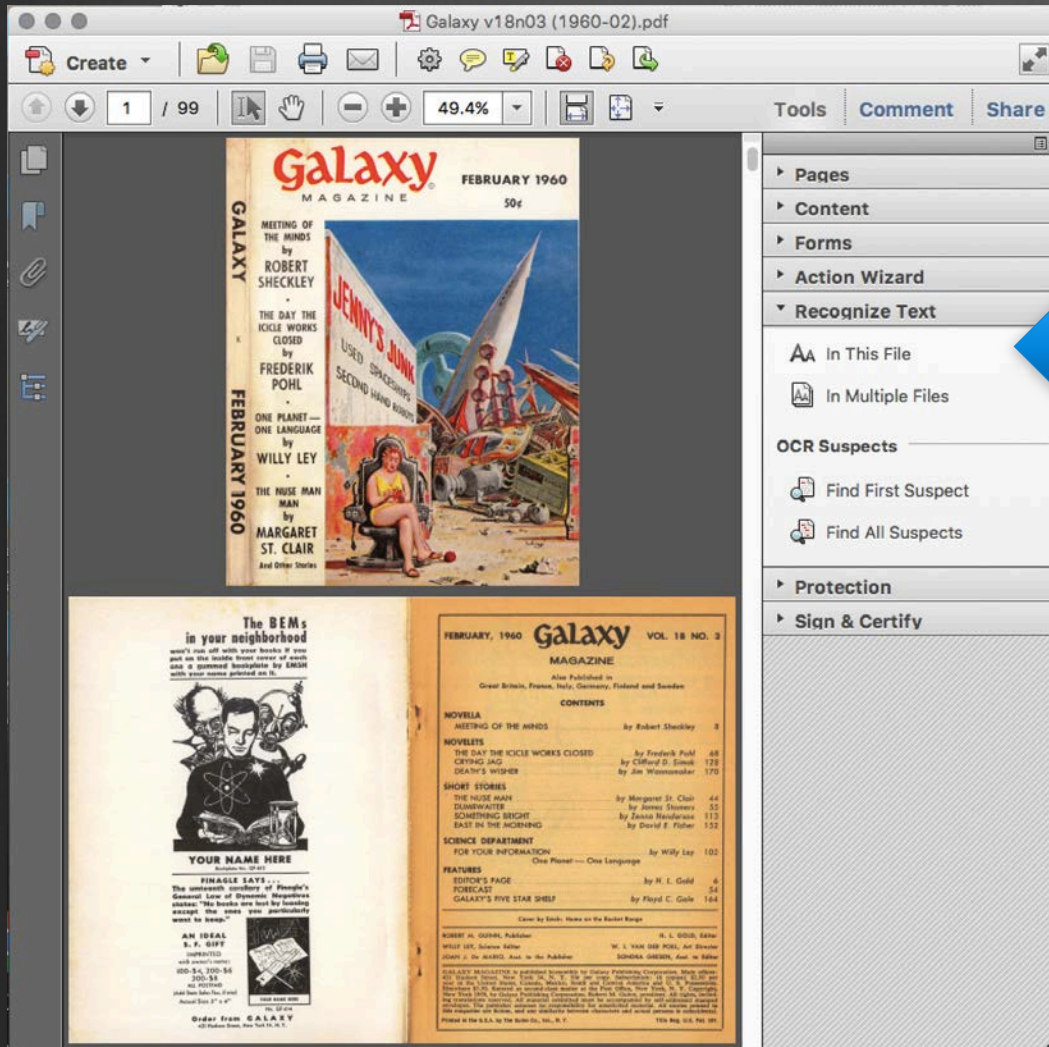
Example: **\$(3997BAA0-329B-42B4-9474-C92E16932862).xxx**

Results **Options**

Loop

Robot Character

Extracting text from PDF pages using Adobe Acrobat Optical Character Recognition (OCR)



This takes quite some time, even with a fast processor and solid-state drive – approximately 80 issues a day

In the early 80s, John Burrows took several years to process and analyse six novels by Jane Austen (Burrows, 1987)

Astounding Stories 1930 v01n01 raw text after OCR

T hey
and trrnJ ie.rs constantly prepared to prove it to the medical pro-
Office trrzveler! .
ets.for :Sore T ..roat
or tr3velin win- sore throat'se.asiestvictims.. Const<Hltly prcserit in eq:n
mouths
are millions of. disease gdrns. The most -
At first sign 6f trot1ble, gqj:gle.witra-. full Listerine and keep it
up sy.
tematically. Also., consult your doctor. LiS... terine colds and sor.t
throat.because'
terare
:tomnion arc those of throat a!<li .it destroys the gt.'rt1ls tk.'at
cause them.
infilll:'nz.l. . . 1 4. .
Though safe to use full
\\'h'-'n, for :un reason, boll\ n:si:.,.ta:l;ce .is strength, it JS fatal. to
even . n;tture. 110 longl'r is. ab(e. to w'ith- the Staphyl<>coccus Aureus
(pus) and
their att:H::ks. ...
qd: t e

Stripping out OCR artefacts

```
# Processes an OCR output file
def process_ocr_output(inputFilename, outputFilename):

    # Read the raw OCR output from the input file
    raw = file_get_contents(inputFilename)

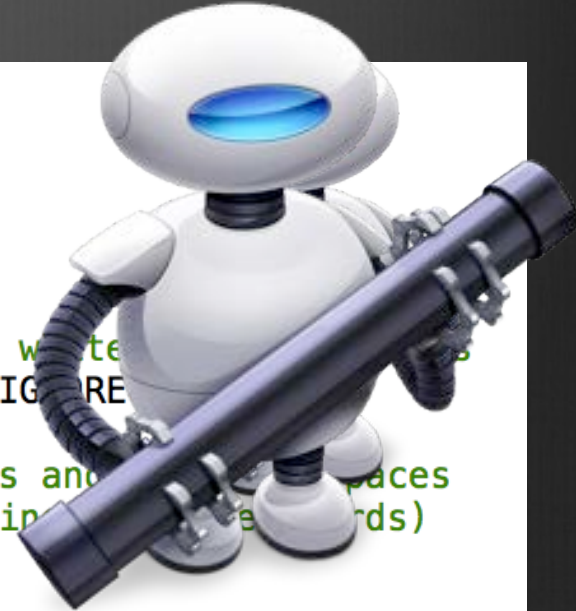
    # Strip any characters that are not alpha-numeric or whitespace
    processed = re.sub('[^a-z0-9\s]', '', raw, flags=re.IGNORECASE)

    # Convert all whitespace characters (such as newlines and spaces)
    # (This ensures that we preserve spaces between newlines and words)
    processed = re.sub('[\s]', ' ', processed)

    # Remove any non-numeric single-character words, except for 'I' or 'a'
    processed = re.sub('[^0-9ai]', '', processed, flags=re.IGNORECASE)

    # Strip excess whitespace
    while (processed.find('  ') != -1):
        processed = processed.replace('  ', ' ')

    # Write the processed data to the output file
    file_put_contents(outputFilename, processed)
```



Courtesy of: Adam Rehn JCU School of Business (Information Technology)

Astounding Stories 1930 v01n01

text after processing

traveler doctor sore road try of traveling with sore throats disease victims constantly
prcserit in en mouths are millions of disease guns the most at first sign a trouble
ggjglewitra full listening and keep it up by tematically also consult your doctor lis
uterine cold and sort throatbecause trade tomnlon arc those of throat ali it destroys
the gtrtlls that cause them infllllnzl a a though safe to use full in for un reason
ball nsitalce is strength it is fatal to even nature 110 longer is abe to with the
staphylococcus aureus pus and their attacks go upper 2000000000 of rangmgto second
bacillus typhosus typhoid in counts handcaus111g disease face coiliitions that weaken
had resist session and government poorairsudden keepabottleoflisterinehandinhome
changes peratureolrexposure and and office tuck one in your bag when could not make
this statement unless contact it people traveling alert coin any kills a ocoooooo
from crowded cars trains and houses st louis fo a the fntistplic st a sect do ti on
man covered hidden possible never vitriol hise his ll income 80801 an luuminatiog
message of vital interest to very au who ever yes to earn my a head see for yours of
roar able volume that he men the fivfiguresalary coat by a a enable any thou years on
the roadwithout collar from a to fo piece every man we were a a copy for to sales of
belllltate it now free them to work for you no what your opinion is now the key to
later saletanshit wm iii to a new fnslht into this fasclnatlog highly paid profession
you tile coupon now a national salesman raining associate tin kept a792 a old chicago
iii us und pnnfples salesmanslj how all can put a aifiohfrug teach jlottotll of til
filter cystic giveyou outfits of radio parts to go ev6fi had low like graduate edward
stand chief operator of war or frank jones 922 guadaluw st san angelo ten builder and
operator of station igni and manager or the best equipped radio shop in tile
southwest or help you a badly business of your own like richard butler he no with st
pua pa who made around 500 a month covered with a job as motorway allen he enrolled
the famous course that pays for tell spare time earnings are easy in radio almost
from the time rou knoll page 1807 must ave nashville tenn made in his spare time

RegEx assistance from Bruce Belson and autocorrect script by Peter Norveig

Summary

(Some) Digital Methods for Text Analysis

- ⦿ Quantitative analysis of content:
 - ⦿ Bulk extraction of text from approximately 50% of all pulp science fiction published in the 20th century – 4,000 volumes, or nearly 200,000,000 words...
 - ⦿ Compiling PDFs, OCR: scripting, automation and RegEx
- ⦿ Text analysis – word frequencies, comparing content across time periods to highlight points of interest
 - ⦿ Using NVivo (which creaks) and R libraries
- ⦿ Benchmarking of results against broader corpus of English text using Google Ngrams

Limitations and caveats

- ⊗ Sources restricted to US and UK periodicals from the mid-1930s until the end of the century – earlier literature tends to be based on crime and western pulp formulas
- ⊗ It's not just about text analysis – editors, agents, writers and publishers also played a part
- ⊗ Online survey data is subject to known factors such as technology use, gender bias of survey responses and survey distribution patterns
- ⊗ Comprehending data on this scale can be challenging

Visualising frequency results

Wordclouds: Stand on Zanzibar (1968)

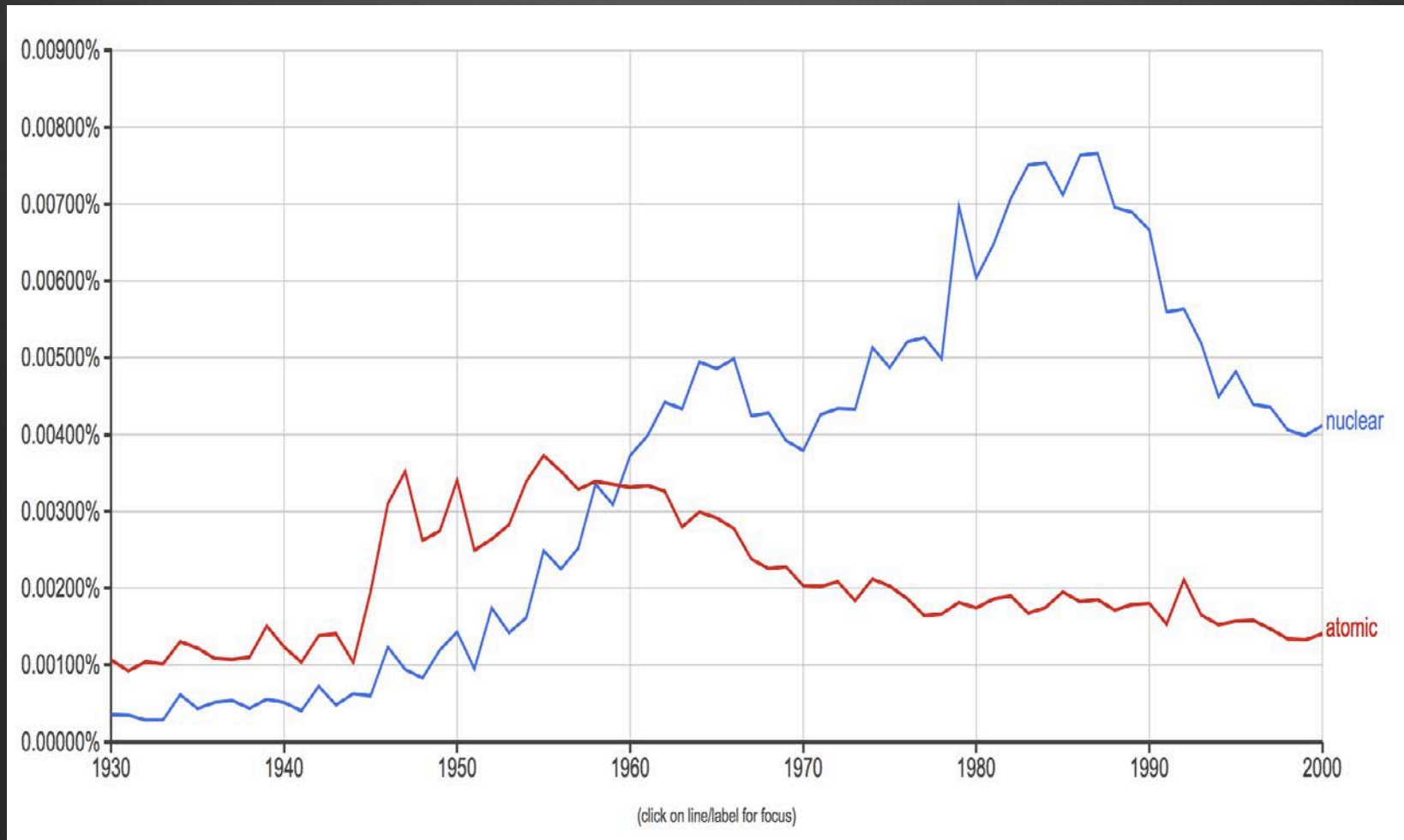


Benchmarking word frequencies 1945 to 1994: Google 'lots of books' vs *Astounding Science Fiction*

	'45 to '54	'55 to '64	'65 to '74	'75 to '84	'85 to '94		'45 to '54	'55 to '64	'65 to '74	'75 to '84	'85 to '94
1	one	one	one	one	one		space	man	man	man	time
2	time	time	new	new	new		man	space	world	space	space
3	new	new	time	time	time		time	time	time	war	man
4	man	man	man	life	way		new	one	space	new	future
5	life	life	life	way	life		stars	new	stars	future	life
6	way	way	way	man	world		world	stars	star	time	mars
7	war	world	world	world	man		earth	war	earth	world	moon
8	world	war	war	war	war		mars	way	life	one	last
9	last	last	last	last	last		moon	world	one	last	planet
10	future	future	future	future	future		one	life	war	world	new
11	earth	space	space	space	space		star	earth	mars	life	way
12	space	earth	earth	earth	earth		way	planet	planet	moon	star
13	star	moon	star	star	star		last	last	way	mars	earth
14	stars	star	moon	moon	moon		war	star	moon	planet	war
15	moon	stars	stars	stars	stars		warfare	moon	future	star	one
16	mars	planet	planet	planet	planet		life	mars	new	earth	
17	planet	mars	mars	mars	mars		planet		warfare	stars	
18									last	way	
Relative Frequency Rank (<i>Ngrams</i>) 1945 to 1994							Relative Frequency Rank <i>Astounding SF</i> 1945 to 1994				

Frequency graphs

Google Ngrams: Atomic vs Nuclear



(Google Ngrams, 2015)

Historical context of word frequency

Ngrams: mapping to historical events

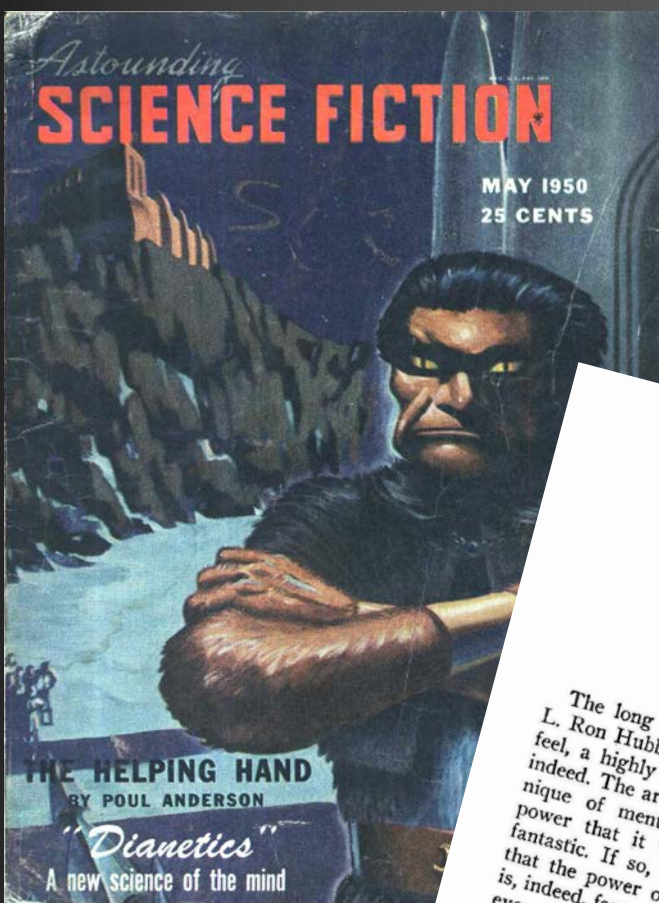


(Google Ngrams, 2015)

Outward reading: “zooming in”

- ⦿ “Outward reading”* of texts including themes and ideas identified in quantitative analysis which are not otherwise readily accessible:
 - ⦿ Editorials,
 - ⦿ Reader feedback surveys (quantitative and qualitative),
 - ⦿ Letters to the editor and advertising
- ⦿ A literary-critical approach

*(Scholes & Wolfman, 2010)



Searching for the obscure: Dianetics (Scientology)

CONCERNING DIANETICS

The long article on *dianetics* by L. Ron Hubbard, in this issue, is, indeed, a highly important publication. The article describes a technique of mental therapy of such power that it will, I know, seem fantastic. If so, it can also be said that the power of the human mind is, indeed, fantastic. I want to assure every reader, most positively and unequivocally, that this article is *not* a hoax, joke, or anything but a direct, clear statement of a *scientific* thesis.

Dr. Joseph A. Winter, M.D., was asked to write the introductory note; he has studied the techniques in detail, and learned the techniques himself. I have investigated the material myself. Dr. Winter can speak as a medical expert; I can only say that my investigations have led me to the conviction that the phenomena I have observed definitely merit publication of this material for wider analysis and testing.

It would have been wholly unfair to publish such revolutionary material until such time as Hubbard's textbook on the technique was available. That book—

Modern Science of Mental Health"—is available now from Hermitage House. It contains the exact description of the procedures used. Hubbard, as an engineer, has tackled the problem of the mind from the scientific method. Basically, that method is:

1. Gather all available data that is, or appears to be, relevant.
2. From the data, form an hypothesis.
3. On the basis of the hypothesis, make a prediction.
4. Experiment to check the validity of the prediction.
5. Vary the experiments, and collect more data.
6. When the new theory breaks down, take the now-collected data and formulate a new hypothesis.
7. Go back to step 3.

Most readers of this magazine are fully accustomed to that method; to those who are not, the above described process is *not* circular. Each turn of the spiral is in more and more detail.

Astounding Science Fiction
v45n03 (1950-05)

NVivo

Open Get Info Edit Paste Merge Format Paragraph Styles Select

Item Clipboard

SOURCES

- Internals
- Externals
- Memos

NODES

- Nodes
- Cases
- Node Matrices

CLASSIFICATIONS

- Source Classifications
- Case Classifications

COLLECTIONS

- Sets
- Memo Links
- Annotations

QUERIES

- Queries
- Results

MAPS

Name	Sources	Referen...	Created On	Created...	Modifi
Dianetics Scientology	82	409	24 Jun 2016, 12:59	CB	24 J

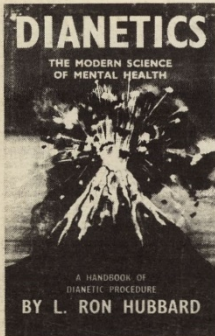
Dianetics Scientology

Summary Reference

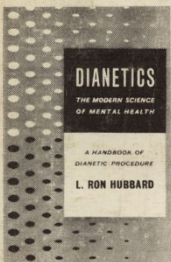
Source Name	In Folder	References	Coverage
Asimovs 2007 v31n10-11 (Oct-Nov 2007)	Internals	2	0.01%
Asimovs 2009 v33n10-11 (Oct-Nov 2009)	Internals	1	0.01%
Asimovs 2011 v35n10-11 (Oct-Nov 2011)	Internals	2	0.01%
Astounding 1949 v44n04 1949_12	Internals	1	0.01%
Astounding v45n01 (1950-03)	Internals	1	0.01%
Astounding v45n02 (1950-04)	Internals	1	0.01%
Astounding v45n03 (1950-05)	Internals	72	0.19%
Astounding v45n04 (1950-06)	Internals	1	0.01%
Astounding v45n05 (1950-07)	Internals	6	0.02%
Astounding v45n06 (1950-08)	Internals	73	0.20%
Astounding v46n01 (1950-09)	Internals	6	0.02%

Whatever happened to

DIANETICS?



ABOVE: the 19th edition of 'Dianetics: the Modern Science of Mental Health' by L. Ron Hubbard, 428 pgs. \$5.00 at most book stores and local Scientology Churches. Send mail orders with coupon BELOW: many are familiar with original book jacket.



DIANETICS IN 1950

Dianetics was first announced to the public in a popular science fiction magazine in 1950. Without formal training or an organization, the superior technology in the first book made Dianetics tremendously popular. Informal groups mushroomed all over the country and hundreds of thousands actively applied Dianetic techniques with amazing results.

Mental upset and psychosomatic ills vanished and miracles were commonplace.

L. Ron Hubbard continued his research to improve the already excellent results from Dianetic technology and went far beyond. He created training methods far superior to existing systems and founded an organization to service the huge demand for the first truly workable technology of the human mind.

Finally in 1966 he completed his research on the full route to the state of CLEAR and TOTAL FREEDOM.

Clip the coupon on the right and mail your name and address for more information sent you at no obligation. Don't delay, do it now.

DIANETICS TODAY

Today we have Scientology (study of knowledge) which was the inevitable result of L. Ron Hubbard's advanced research into the mind. Scientology goes beyond the mind into the realm of the human spirit.

Dianetics was only part of the answer and now, with Scientology techniques, Dianetics is 100% workable.

Dianetics takes a person up to a well and happy human being. Scientology goes beyond to higher states of existence. Scientology is far more popular than Dianetics ever was alone.

Today hundreds of Scientology Centers and organizations circle the globe and Scientologists number in the millions. Scientology doubles every year.

Get all the answers! Send in your name and address for more information about Modern Dianetics, Scientology and our 20th Anniversary Convention.

Get more exciting, vital information about Dianetics, Scientology and the Grand National Convention. Send your name and address to:

DEPARTMENT I
Church of Scientology
916 South Westlake
Los Angeles, Calif. 90006

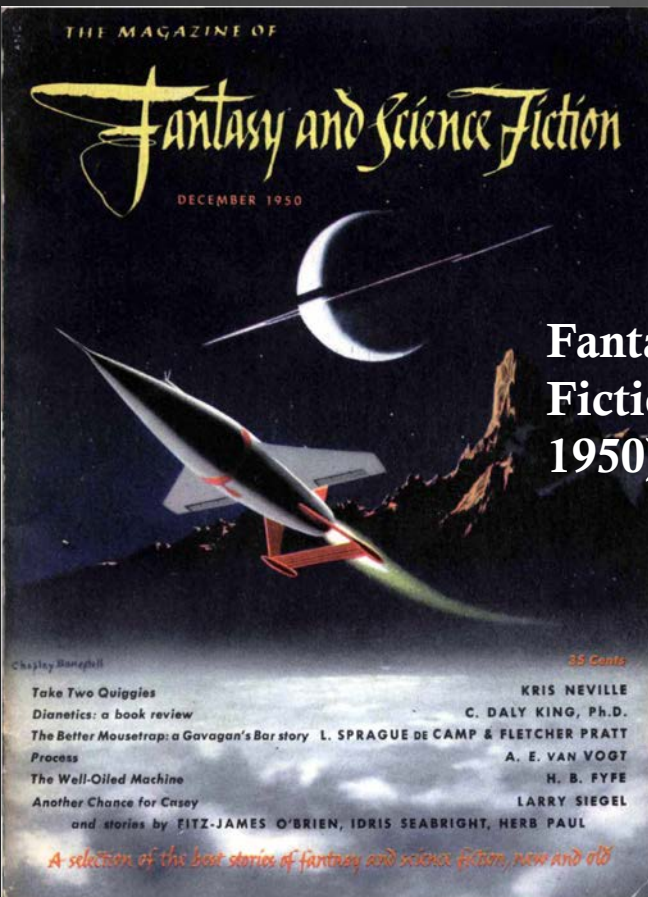
Name
Address
State Zip
☐ Enclosed is \$5 for a copy of Dianetics:
the Modern Science of Mental Health.

L. Ron Hubbard presents Scientology's
GRAND NATIONAL CONVENTION OF THE 20TH
ANNIVERSARY OF DIANETICS & TOTAL FREEDOM
July 3rd, 4th and 5th at the west's most beautiful and modern convention center, The LONG BEACH ARENA, Long Beach California.
Here is the complete story of the tremendous growth of Dianetics and its evolution into Scientology, its amazing technology and the culmination of 20 years of "miracles beyond belief." Plan now to attend this event.

Advertising

There are over 100
Dianetics /
Scientology adverts
from 1950 to the
present day -
frequency and
content is variable

Worlds of Tomorrow
v05n01 p164 Dianetics
Advertising

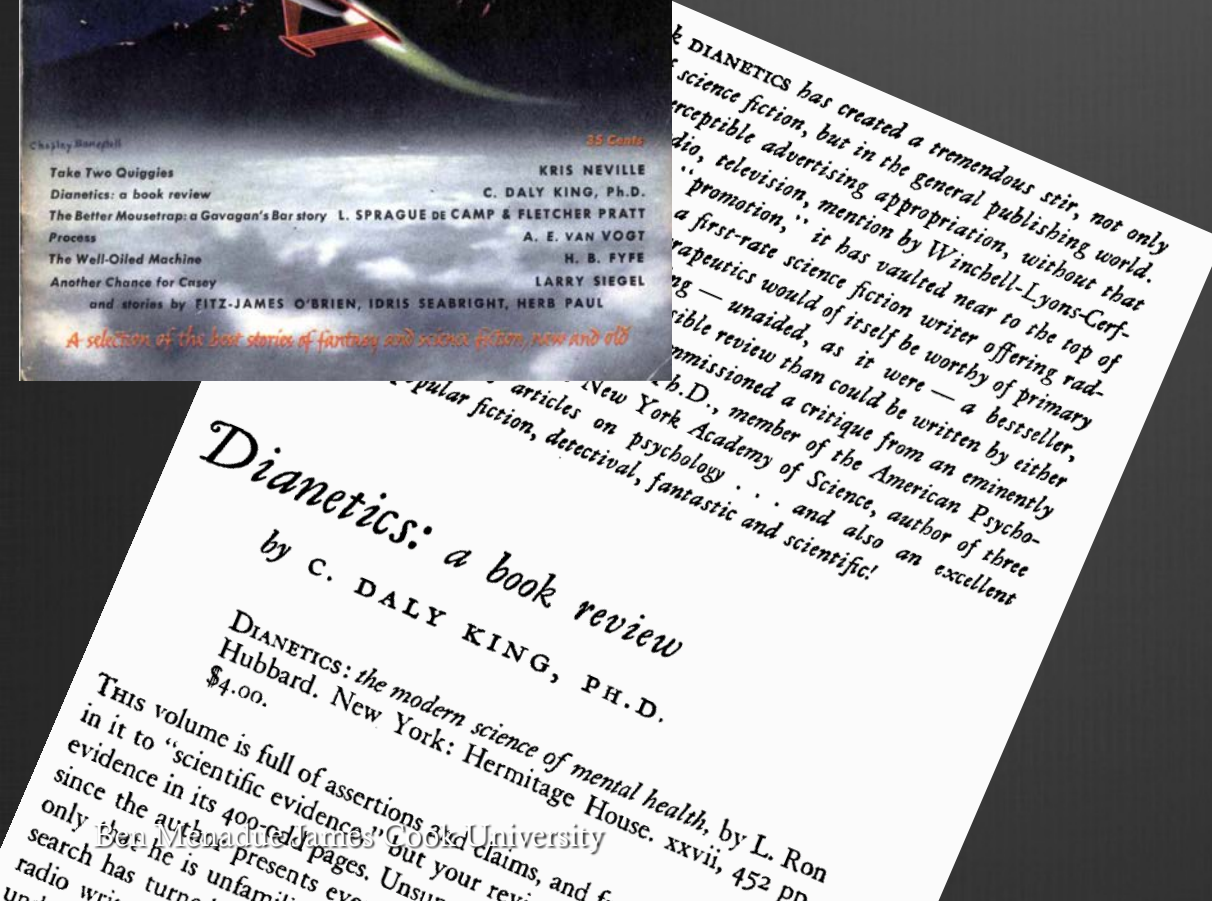


Discovering “Lost” literature

Fantasy and Science
Fiction v01n01 (December
1950)

“Vianetics: a hook review
by C. DALY KING, PH.D.
DIANETICS: the modern
science of mental health, by L.
Ron Hubbard. New York:
Hermitage House. xxvii, 452
pp. \$4.00.

THIS volume is full of
assertions and claims, and
frequent reference is made in it
to “scientific evidence,” but
your reviewer could find no
item of such evidence in its
400-odd pages.”



Outputs

- ⦿ Synthesis of the Qualitative with the Quantitative to identify correlations and test validity
- ⦿ Investigation of the world view: enduring human themes and mythologies
- ⦿ Reading the spirit of the age: how people responded to content that is a specific product of their time – new and revolutionary ideas, influence of historical events and changes in popular culture and society

Questions



Yes, I'd like to hear it, Hal. Sing it for me.

References

References cited in the text

Brunner (1968), 'Stand on Zanzibar', Del Rey, NY.

Burrows, J. F. (1987). *Computation into criticism: a study of Jane Austen's novels and an experiment in method*. Oxford: Oxford University Press.

Freedman, Carl (2000). *Critical Theory and Science Fiction*. Wesleyan University Press, Connecticut.

Hubbard, L. Ron (1950) 'Dianetics, The Evolution of a Science', ASF May 1950, Street & Smith, NYC

King, C. D. (1950). Dianetics: a book review. *The Magazine of Fantasy and Science Fiction*, 1(5), 99.

Menadue, C. B. (2015). *Science fiction and fantasy experience survey*. Retrieved from:

<https://researchdata.ands.org.au/science-fiction-fantasy-experience-survey>

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books, T., . . . Aiden, E. L. (2011).

Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014), 176-182.

Nordlund, S (2011) *J.G. Ballard – Future Now* retrieved from: https://www.youtube.com/watch?v=SS6MWpFX_N0

QSR International. (2016). Nvivo (Version 11.4.1). Doncaster, VIC: QSR International. Retrieved from

<http://www.qsrinternational.com/>

Scholes, R., & Wulfman, C. (2010). *Modernism in the Magazines: An Introduction*: Yale University Press.

Van Vogt, Alfred Eaton (1939), 'Black Destroyer', *Astounding Science Fiction*, Street & Smith, NY, July 1939

Images

Images used are published under creative commons, fair use or copyright expired conditions unless otherwise noted below:

Space photography: NASA; public domain.

Astounding Science Fiction Covers, © Street & Smith Publications.

Book cover images all © Chris Foss www.chrisfossart.com/ except cover to 2001 *A Space Odyssey* and associated animated gif images, © Metro Goldwyn Meyer (1968)

Wordle (2015) <http://www.wordle.net/create>